**Rare and Common Variants Conferring Risk of Tooth Agenesis**

L. Jonsson , T.E. Magnusson, A. Thordarson, T. Jonsson, F. Geller, B. Feenstra, M. Melbye, E.A. Nohr, S. Vucic, B. Dhamo, F. Rivadeneira, E.M. Ongkosuwito, E.B. Wolvius, E.J. Leslie, M.L. Marazita, B.J. Howe, L.M. Moreno Uribe, I. Alonso, M. Santos, T. Pinho, R. Jonsson, G. Audolfsson, L. Gudmundsson, M.S. Nawaz, S. Olafsson, O. Gustafsson, A. Ingason, U. Unnsteinsdottir, G. Bjornsdottir, G.B. Walters, M. Zervas, A. Oddsson, D.F. Gudbjartsson, S. Steinberg, H. Stefansson, and K. Stefansson

## Appendix

## Appendix Subjects and Methods

### Study Populations

**Icleandic sample.** In total, the TA cases were missing 5,423 teeth with an average of 2.8 missing teeth per subject in the TA population. Of the TA cases, 1,205 subjects were female (62.0 %) and 739 subjects were male (38.0 %). The prevalence of missing teeth at specific dental positions are shown in Fig 2. The most commonly missing teeth are the mandibular second premolars ($N$ = 1,987), followed by the maxillary second premolars ($N$ = 1,034) and the maxillary lateral incisors ($N$ = 998). In the association analysis between TA and known OFC variants, 22 subjects with OFC were excluded from the analyses.

**Danish TA follow-up sample.** The Danish replication study was based on 4,774 individuals from the Danish National Birth Cohort, who were genotyped on Illumina bead chips in GWAS of preterm birth and obesity, as described earlier[1]. We imputed unobserved genotypes based on phased haplotypes from the integrated Phase I release of the 1000 Genomes Project using the software programs SHAPEIT[2] and IMPUTE2[3].

Dental data for all Danish individuals were retrieved from the nationwide dental registry for children, SCOR, which was established in 1972 alongside the initiation of free municipal dental services to Danish children and adolescents from birth to the age of 18 years[4]. We combined all observations on erupted permanent teeth after age 6. Controls ($N$ = 4,667) were defined as individuals with all 28

permanent teeth (excluding third molars) erupted by age 14, whereas cases ($N$ = 107) were defined as individuals with consistently less than 28 permanent teeth (excluding third molars) and at least one record after their 16$^{th}$ birthday. Information on the not erupted teeth was available and we were able to perform a subgroup analysis of mandibular second premolars based on 66 cases, for maxillary lateral incisors the low number of 14 cases precluded any meaningful analyses.

We used logistic regression to analyze all imputed variants, testing for differences in allele dosages between cases and controls under an additive genetic model as implemented in SNPTEST[5]. The modest inflation of the test statistic was adjusted for by applying genomic control[6] ($\lambda$ = 1.009 for both general TA and mandibular second premolars). The study protocol was approved by the Danish Scientific Ethical Committee and the Danish Data Protection Agency for all subjects.

**Dutch TA follow-up sample.** TA was assessed by a dentist from dental panoramic radiographs of the children (mean age = 9.8 ± 0.3 years) participating in the Generation R Study, a population-based cohort study from fetal life until adulthood, established in Rotterdam, the Netherlands[7]. The study protocol was approved by the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam (MEC 217.595/2002/20). In total, 150 children were diagnosed with TA of one or more teeth (males = 73 and females = 77) in the cohort, with the remaining 2,696 used as controls (males = 1,349 and females = 1,347). Of the 150 children, 82 subjects were missing the mandibular second premolars and 32 subjects were missing the maxillary lateral incisors.

Genetic data have been generated by a genome-wide association scan (GWAS) using Illumina HumanHap 610 or 660 Quad chips (Illumina Inc., San Diego, USA). Genotypes were imputed to the 1000 Genomes Project reference panel (March 2013 version). The GWAS and imputed datasets underwent a stringent QC process, which is described in detail in the study by Medina-Gomez, *et al.*[8]. Individuals of European ($N$ = 1,624) and non-European ($N$ = 1,222) genetic ancestry were grouped together in the analysis.

Association between TA and GWAS SNPs was carried out using a logistic regression framework adjusting for age, sex, and population stratification (10 genomic principal components) in the Generation R cohort using MACH2DAT as implemented in GRIMP[9]. The genomic inflation factors (λ) for the general TA, maxillary lateral incisors and mandibular second premolars were 1.016, 1.048 and 1.010, respectively.

**TA follow-up sample from Pittsburgh, USA.** A total of 4,151 subjects were recruited from multiple sites in the United States, Guatemala, Hungary, Colombia, Argentina, and the Philippines. The sample included 768 individuals with a nonsyndromic cleft lip with or without cleft palate, 1,854 relatives without a cleft, and 1,529 control individuals with no history of orofacial clefting. 2,551 subjects were over the age of 18; 522 were between 12 and 18; 1,076 subjects were under the age of 12. Questionnaires recording dental history were collected along with in-person dental exams or intraoral photos. Intraoral photos were evaluated by a single rater, calibrated against experienced dentists. Each tooth was evaluated and determined to be missing due to agenesis based on the status of the dentition[10]. The phenotypes for analysis were (1) agenesis of the maxillary lateral incisors or (2) any agenesis, excluding the 2$^{nd}$ and 3$^{rd}$ molars.

Samples were genotyped on the Illumina HumanCore+Exome array, phased with SHAPEIT2, and imputed to the 1000 Genomes Phase 3 reference panel. Imputed genotype probabilities were converted to most-likely genotypes using GTOOL, keeping only genotypes if the probability was >0.9. All SNPs considered for replication had INFO scores greater than 0.8.

Statistical analysis was performed using EMMAX (Efficient Mixed-Model Association eXpedited), a linear mixed-model that accounted for sample structure (relatedness and population structure) with a kinship matrix. An association with TA was tested in individuals without orofacial clefting (combining controls and family members of orofacial clefting cases). The sample was also stratified into genetically-determined European and non-European ancestry groups.

The sample included in total 147 TA cases and 3,236 controls. The statistical association method used in this sample with related subjects only made it possible to only include direction of effect in the anlyses. In this study, the European TA sample of subjects without orofacial clefting (44 TA cases and 1,352 controls) were included.

**TA follow-up sample from Portugal.** The Portuguese sample included 102 individuals with agenesis of the maxillary lateral incisors and 204 controls (without any kind of hypodontia)[11]. All subjects in this study were observed by experienced clinicians and TA was confirmed radiographically. The study followed the "Strengthening the Reporting of Observational Studies in Epidemiology" guidelines[12] and informed consent was obtained from all participants. Three variants (rs35956082, rs7552 and rs9825432) were genotyped in the Portuguese sample by PCR and Sanger sequencing. Association analysis between agenesis of maxillary lateral incisors and the three SNPs were carried out using logistic regression under an additive genetic model adjusting for gender. All statistical analyses were performed using R software.

**Variants associated with TA-related phenotypes tested in the Icelandic sample**

We tested variants previously associated at a genome wide significance threshold with OFC or timing of tooth eruption in the Icelandic TA sample, including the samples for agenesis of the specific teeth maxillary lateral incisors, mandibular second premolars and maxillary second premolars.

In total, 40 OFC variants were tested in our TA data. Twenty previously reported OFC variants were chosen based on the study by Ludwig, *et al.*[13], in which the most credible SNP in European samples was chosen at each of the 20 significant OFC loci. In addition, 20 novel OFC variants identified after that study were also chosen[13-15]. At each locus, SNPs with $r^2 > 0.2$ (based on the Icelandic data set) with a more significant SNP were discarded; thus, a single SNP was tested for each association signal. Variants that were significantly associated with TA after Bonferroni correction ($P = 0.05/160 = 0.00031$, correcting

for 40 variants and 4 phenotypes) and that had the same risk allele for TA and OFC were included for replication efforts.

For timing of primary and permanent tooth eruption, we tested variants at 17 previously reported loci[1,16,17] in our TA sample. Variants having $r^2 > 0.2$ (based on the Icelandic data set) with a more significant variant were not included. Variants surviving Bonferroni correction for 17 variants and four TA phenotypes ($P_{thresh} = 0.05/68 = 0.00074$) were regarded as significant associations.

**Gene-set enrichment analysis (GSEA)**

We performed GSEA for the TA-associated markers (N = 11) using GO database through INRICH as described in online methods for INRICH[18]. In this analysis, a total of 11,919 GO terms mapping to 19,432 genes were tested. For GSEA, we tested associated intervals belonging to sequence variants using their tight LD ($r^2 > 0.8$) prioritizing strongest cis-eQTL in region (i.e. in ±250 KB of each sequence variant). Moreover, to include gene boundaries, we used a flanking distance of ±100 KB from associated intervals.

# Appendix Tables

**Appendix Table 1.** Sequence variants associated with orofacial cleft (OFC) tested for association with TA and selective TA*

| | | | | | Tooth Agenesis (*N* = 1,944) | | Maxillary Lateral Incisors (*N* = 600) | | Mandibular Second Premolars (*N* = 1,196) | | Maxillary Second Premolars (*N* = 600) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| marker[†] | chr | Position (hg38) | EAF (%) | EA/OA | OR (95% CI) | *P* | OR (95% CI) | *P* | OR (95% CI) | *P* | OR (95% CI) | *P* |
| rs4920524 | 1 | 18,651,878 | 42.3 | G/T | 1.00 (1.00, 1.00) | 0.97 | 0.94 (0.79, 1.12) | 0.49 | 1.05 (0.93, 1.19) | 0.45 | 1.06 (0.91, 1.24) | 0.46 |
| rs35298667 | 1 | 94,075,119 | 19.0 | T/C | 0.99 (0.90, 1.09) | 0.84 | 0.99 (0.86, 1.15) | 0.89 | 0.98 (0.84, 1.14) | 0.80 | 0.95 (0.77, 1.17) | 0.62 |
| rs658860 | 1 | 209,817,204 | 19.0 | C/T | 0.96 (0.86, 1.07) | 0.45 | 0.86 (0.69, 1.07) | 0.18 | 0.95 (0.82, 1.10) | 0.49 | 1.05 (0.87, 1.27) | 0.61 |
| rs287982 | 2 | 9,832,313 | 25.0 | C/T | 1.05 (0.95, 1.17) | 0.36 | 0.89 (0.75, 1.06) | 0.20 | 1.06 (0.93, 1.21) | 0.39 | 1.15 (0.97, 1.37) | 0.11 |
| rs5829552[‡] | 2 | 16,545,695 | 22.5 | TA/T | 1.07 (0.97, 1.18) | 0.19 | 1.55 (1.30, 1.85) | **$1.4 \times 10^{-6}$** | 0.94 (0.83, 1.07) | 0.35 | 1.04 (0.85, 1.27) | 0.70 |
| rs6740960 | 2 | 41,954,539 | 43.7 | A/T | 1.06 (0.97, 1.16) | 0.22 | 0.92 (0.78, 1.09) | 0.34 | 1.10 (0.98, 1.24) | 0.11 | 1.02 (0.89, 1.17) | 0.78 |
| rs7590268 | 2 | 43,312,986 | 21.8 | G/T | 1.02 (0.92, 1.14) | 0.72 | 1.07 (0.89, 1.29) | 0.47 | 0.98 (0.84, 1.15) | 0.80 | 1.05 (0.88, 1.26) | 0.60 |
| rs7632427[§] | 3 | 89,485,227 | 40.1 | C/T | 0.97 (0.90, 1.05) | 0.44 | 0.96 (0.81, 1.14) | 0.65 | 1.02 (0.91, 1.14) | 0.73 | 0.87 (0.74, 1.03) | 0.098 |
| rs9832134 | 3 | 100,117,878 | 40.9 | C/T | 1.01 (0.93, 1.10) | 0.81 | 0.97 (0.84, 1.12) | 0.67 | 1.01 (0.86, 1.18) | 0.90 | 1.01 (0.74, 1.37) | 0.95 |
| rs76479869 | 3 | 189,835,583 | 6.5 | T/C | 0.81 (0.67, 0.98) | 0.03 | 0.94 (0.68, 1.30) | 0.71 | 0.78 (0.61, 1.00) | 0.048 | 0.70 (0.49, 1.01) | 0.056 |
| rs1907989 | 4 | 4,817,198 | 40.8 | G/A | 0.99 (0.87, 1.13) | 0.88 | 0.96 (0.82, 1.13) | 0.62 | 1.03 (0.90, 1.18) | 0.67 | 1.02 (0.89, 1.17) | 0.78 |
| rs908822 | 4 | 123,985,102 | 6.2 | T/C | 1.06 (0.88, 1.27) | 0.53 | 1.07 (0.77, 1.49) | 0.69 | 1.01 (0.83, 1.22) | 0.92 | 1.19 (0.86, 1.64) | 0.29 |
| rs10462065 | 5 | 44,068,744 | 12.9 | A/C | 0.95 (0.83, 1.09) | 0.46 | 1.00 (1.00, 1.00) | 0.98 | 0.95 (0.81, 1.11) | 0.52 | 0.96 (0.75, 1.23) | 0.74 |
| rs9381107 | 6 | 9,469,005 | 15.5 | A/G | 0.97 (0.87, 1.08) | 0.59 | 0.83 (0.66, 1.05) | 0.11 | 0.97 (0.85, 1.11) | 0.66 | 1.03 (0.83, 1.28) | 0.79 |
| rs13317 | 8 | 38,411,996 | 28.0 | C/T | 1.02 (0.89, 1.16) | 0.77 | 1.11 (0.94, 1.32) | 0.23 | 0.96 (0.86, 1.08) | 0.49 | 0.88 (0.74, 1.05) | 0.16 |
| rs12543318 | 8 | 87,856,112 | 30.4 | C/A | 0.93 (0.85, 1.02) | 0.13 | 0.91 (0.76, 1.09) | 0.32 | 0.88 (0.78, 1.00) | 0.047 | 0.80 (0.67, 0.96) | 0.014 |
| rs957448 | 8 | 94,529,074 | 27.1 | G/A | 1.07 (0.97, 1.18) | 0.16 | 1.02 (0.86, 1.22) | 0.82 | 1.08 (0.95, 1.22) | 0.23 | 1.15 (0.97, 1.37) | 0.11 |
| rs72728755 | 8 | 128,978,136 | 18.6 | A/T | 1.15 (1.03, 1.29) | 0.016 | 1.24 (1.02, 1.51) | 0.033 | 1.15 (1.00, 1.32) | 0.049 | 0.99 (0.78, 1.26) | 0.94 |
| rs10512248 | 9 | 95,497,421 | 38.1 | G/T | 1.16 (1.06, 1.27) | 0.0015 | 1.07 (0.91, 1.26) | 0.41 | 1.24 (1.11, 1.39) | 0.0002 | 1.31 (1.12, 1.54) | 0.0010 |
| rs4582663 | 9 | 97,867,545 | 36.0 | T/G | 0.99 (0.92, 1.07) | 0.80 | 0.78 (0.66, 0.92) | 0.0040 | 1.06 (0.94, 1.20) | 0.34 | 1.21 (1.03, 1.42) | 0.020 |
| rs7092957 | 10 | 117,041,514 | 16.6 | G/A | 0.97 (0.87, 1.08) | 0.59 | 0.85 (0.68, 1.06) | 0.15 | 1.06 (0.90, 1.24) | 0.47 | 0.98 (0.79, 1.22) | 0.85 |
| rs3741442 | 12 | 52,952,966 | 0.9 | T/C | 0.85 (0.52, 1.39) | 0.52 | 0.73 (0.25, 2.11) | 0.56 | 0.63 (0.32, 1.23) | 0.18 | 0.88 (0.40, 1.94) | 0.75 |
| rs705704 | 12 | 56,041,628 | 34.5 | A/G | 0.95 (0.86, 1.05) | 0.32 | 0.93 (0.78, 1.11) | 0.42 | 0.94 (0.84, 1.05) | 0.29 | 1.01 (0.90, 1.14) | 0.87 |
| rs2304269 | 12 | 71,686,492 | 5.3 | C/T | 0.96 (0.80, 1.15) | 0.66 | 0.92 (0.64, 1.33) | 0.66 | 0.98 (0.75, 1.28) | 0.88 | 0.97 (0.64, 1.47) | 0.89 |
| rs7999259 | 13 | 80,123,426 | 23.2 | A/G | 1.02 (0.93, 1.12) | 0.69 | 0.96 (0.79, 1.16) | 0.68 | 1.00 (1.00, 1.00) | 0.95 | 0.99 (0.87, 1.13) | 0.88 |
| rs7148069 | 14 | 51,372,927 | 31.5 | T/C | 0.96 (0.88, 1.05) | 0.38 | 0.90 (0.75, 1.08) | 0.26 | 0.98 (0.88, 1.09) | 0.72 | 1.00 (1.00, 1.00) | 0.97 |
| rs60454187 | 14 | 51,389,848 | 36.9 | C/G | 1.08 (0.99, 1.18) | 0.10 | 1.20 (1.02, 1.41) | 0.027 | 1.10 (0.98, 1.24) | 0.12 | 1.07 (0.92, 1.25) | 0.39 |
| rs1243572 | 14 | 94,913,162 | 21.7 | T/C | 1.00 (1.00, 1.00) | 0.92 | 0.92 (0.75, 1.12) | 0.41 | 1.03 (0.89, 1.20) | 0.70 | 1.22 (1.02, 1.47) | 0.034 |
| rs2600520 | 15 | 32,762,318 | 29.6 | T/G | 1.03 (0.93, 1.14) | 0.56 | 1.10 (0.92, 1.31) | 0.28 | 0.99 (0.87, 1.12) | 0.88 | 0.99 (0.77, 1.28) | 0.94 |
| rs4774467 | 15 | 63,019,226 | 28.5 | T/C | 1.09 (0.98, 1.21) | 0.10 | 1.07 (0.89, 1.28) | 0.46 | 1.09 (0.96, 1.23) | 0.17 | 1.14 (0.96, 1.35) | 0.13 |
| rs57490152 | 15 | 74,794,303 | 7.2 | C/- | 1.00 (1.00, 1.00) | 0.99 | 1.02 (0.74, 1.40) | 0.90 | 1.06 (0.86, 1.31) | 0.59 | 0.95 (0.67, 1.34) | 0.77 |
| rs9938468 | 16 | 3,929,044 | 42.4 | C/T | 0.95 (0.87, 1.03) | 0.23 | 0.93 (0.79, 1.10) | 0.39 | 0.91 (0.81, 1.03) | 0.12 | 1.06 (0.91, 1.24) | 0.47 |
| rs58772677 | 17 | 9,016,098 | 37.7 | A/C | 0.93 (0.84, 1.02) | 0.14 | 0.89 (0.76, 1.04) | 0.15 | 0.90 (0.80, 1.01) | 0.078 | 0.93 (0.79, 1.10) | 0.39 |
| rs12944377 | 17 | 9,044,391 | 44.2 | C/T | 1.01 (0.87, 1.17) | 0.90 | 0.99 (0.75, 1.31) | 0.94 | 0.99 (0.89, 1.10) | 0.85 | 0.98 (0.86, 1.12) | 0.77 |
| rs1838105 | 17 | 46,931,569 | 35.5 | A/G | 1.06 (0.96, 1.17) | 0.24 | 1.04 (0.89, 1.21) | 0.62 | 1.07 (0.95, 1.20) | 0.26 | 1.04 (0.89, 1.21) | 0.62 |
| rs227727 | 17 | 56,699,594 | 47.8 | T/A | 0.90 (0.82, 0.99) | 0.027 | 0.87 (0.74, 1.03) | 0.10 | 0.88 (0.79, 0.98) | 0.023 | 0.96 (0.82, 1.13) | 0.61 |
| rs8071332 | 17 | 63,064,592 | 23.5 | G/A | 1.12 (1.01, 1.24) | 0.033 | 1.25 (1.04, 1.50) | 0.015 | 1.08 (0.95, 1.23) | 0.25 | 1.04 (0.86, 1.26) | 0.69 |
| rs146108265 | 19 | 2,051,262 | 26.6 | GCCA/- | 1.07 (0.97, 1.18) | 0.18 | 1.14 (0.94, 1.38) | 0.18 | 1.09 (0.96, 1.23) | 0.17 | 1.07 (0.88, 1.30) | 0.49 |
| rs8113265 | 19 | 32,855,298 | 44.5 | G/A | 1.07 (0.98, 1.17) | 0.15 | 1.05 (0.89, 1.24) | 0.57 | 1.08 (0.97, 1.20) | 0.17 | 1.00 (1.00, 1.00) | 1.00 |
| rs4812450 | 20 | 40,644,319 | 41.3 | G/C | 1.12 (1.03, 1.22) | 0.012 | 1.24 (1.05, 1.46) | 0.011 | 1.05 (0.94, 1.18) | 0.40 | 1.08 (0.93, 1.26) | 0.32 |

OR: Odds Ratio. EA: Effect Allele for which OR is shown. OA: Other Allele . EAF: effect allele frequency (%) in the Icelandic sample. N: Number of TA cases. Bold indicates significant association after Bonferroni corrected $P = 0.05/160 = 0.00031$ (correcting for 40 variants and 4 phenotypes). *Known CLP cases ($N = 22$) were excluded from the analyses in the Icelandic sample. †Twenty previously reported OFC variants were chosen based on the study by Ludwig, *et al.* [13], in which the most credible SNP in European samples were chosen in the 20 OFC loci. In addition, 20 novel OFC variants identified after that study were also chosen[13-15]. ‡The following rs-names are reported for this variant: rs5829552, rs397784935 and rs869032736. §We included rs7632427 instead of rs6772813 from the study by Ludwig, et al. [13] due to low info in imputation in Iceland.

**Appendix Table 2.** Results from association analysis between TA and loci previously associated with timing of primary[16,17] or permanent[1] teeth eruption.

| | | | | | | Tooth Agenesis (*N* = 1,944) | | Maxillary Lateral Incisors (*N* = 600) | | Mandibular Second Premolars (*N* = 1,196) | | Maxillary Second Premolars (*N* = 600) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Marker* | gene | chr | position | EAF (%) | EA/OA | OR | *P* | OR | *P* | OR | *P* | OR | *P* |
| rs2281845[1] | *CACNA1S* | 1 | 201,112,815 | 40.8 | T/C | 1.38 (1.26, 1.51) | **1.1 x $10^{-12}$** | 1.33 (1.13, 1.56) | **0.00050** | 1.45 (1.29, 1.62) | **1.3 x $10^{-10}$** | 1.61 (1.38, 1.88) | **1.9 x $10^{-9}$** |
| rs4491709‡[1,16] | . | 2 | 217,030,033 | 32.4 | C/T | 0.99 (0.87, 1.12) | 0.88 | 0.91 (0.76, 1.08) | 0.29 | 0.94 (0.83, 1.07) | 0.35 | 1.16 (0.98, 1.37) | 0.08 |
| rs6568401[16] | . | 6 | 105,740,943 | 28.2 | C/T | 1.02 (0.91, 1.14) | 0.73 | 0.95 (0.79, 1.14) | 0.59 | 1.02 (0.91, 1.14) | 0.73 | 1.04 (0.88, 1.23) | 0.65 |
| rs1799922[16] | *OPN1SW* | 7 | 128,775,141 | 38.3 | G/T | 0.94 (0.86, 1.03) | 0.20 | 0.86 (0.73, 1.02) | 0.075 | 0.94 (0.83, 1.06) | 0.32 | 0.92 (0.79, 1.07) | 0.29 |
| rs10740993[16] | *CACNB2* | 10 | 18,153,553 | 41.5 | T/C | 0.96 (0.88, 1.05) | 0.36 | 1.02 (0.82, 1.27) | 0.86 | 0.94 (0.85, 1.04) | 0.25 | 0.93 (0.80, 1.09) | 0.36 |
| rs7924176[1,16] | *ADK* | 10 | 74,536,031 | 46.1 | G/A | 0.90 (0.82, 0.99) | 0.026 | 0.85 (0.72, 1.00) | 0.054 | 0.93 (0.83, 1.04) | 0.21 | 0.93 (0.79, 1.10) | 0.40 |
| rs4937076[16] | *CDON* | 11 | 125,956,807 | 45.7 | G/A | 1.05 (0.96, 1.15) | 0.30 | 1.04 (0.87, 1.25) | 0.68 | 1.00 (1.00, 1.00) | 0.96 | 1.15 (0.98, 1.35) | 0.081 |
| rs12229918‡[16,17] | *MSRB3* | 12 | 65,368,278 | 37.9 | C/G | 0.96 (0.87, 1.06) | 0.42 | 0.91 (0.78, 1.07) | 0.25 | 1.01 (0.84, 1.21) | 0.91 | 0.94 (0.80, 1.11) | 0.47 |
| rs12424086‡[1,16] | *HMGA2* | 12 | 65,970,729 | 17.5 | C/T | 1.06 (0.95, 1.18) | 0.29 | 1.19 (0.98, 1.45) | 0.087 | 1.03 (0.89, 1.19) | 0.69 | 0.95 (0.77, 1.18) | 0.64 |
| rs9316505[16] | *DLEU7* | 13 | 50,816,462 | 42.4 | G/A | 1.02 (0.93, 1.12) | 0.69 | 1.01 (0.80, 1.27) | 0.93 | 1.03 (0.92, 1.15) | 0.61 | 0.92 (0.78, 1.08) | 0.32 |
| rs997154[16] | *C14orf93* | 14 | 22,995,273 | 19.5 | A/G | 0.91 (0.81, 1.02) | 0.099 | 0.99 (0.85, 1.16) | 0.90 | 0.87 (0.76, 1.00) | 0.051 | 0.82 (0.67, 1.00) | 0.054 |
| rs17563[16] | *BMP4* | 14 | 53,950,804 | 43.8 | A/G | 1.03 (0.95, 1.12) | 0.49 | 1.04 (0.89, 1.21) | 0.61 | 1.00 (1.00, 1.00) | 0.99 | 1.13 (0.96, 1.33) | 0.13 |
| rs1956529[17] | *RAD51B* | 14 | 68,322,207 | 37.8 | C/T | 1.06 (0.97, 1.16) | 0.20 | 1.07 (0.92, 1.25) | 0.39 | 1.02 (0.90, 1.15) | 0.75 | 1.13 (0.96, 1.32) | 0.13 |
| rs1994969‡[16,17] | *IGF2BP1* | 17 | 49,003,069 | 46.9 | G/T | 0.97 (0.88, 1.07) | 0.52 | 0.97 (0.83, 1.14) | 0.71 | 0.98 (0.87, 1.10) | 0.74 | 1.02 (0.87, 1.19) | 0.80 |
| rs412000[16] | *TEX14* | 17 | 58,631,697 | 44.4 | G/C | 1.12 (1.02, 1.23) | 0.016 | 1.09 (0.93, 1.27) | 0.27 | 1.17 (1.05, 1.31) | 0.0058 | 1.03 (0.85, 1.24) | 0.76 |
| rs8080944‡[16,17] | . | 17 | 70,189,445 | 40.4 | G/A | 0.94 (0.86, 1.02) | 0.15 | 0.94 (0.81, 1.09) | 0.42 | 0.90 (0.81, 1.01) | 0.063 | 0.89 (0.76, 1.05) | 0.16 |
| rs11796357‡[16] | *EDA* | X | 69,578,860 | 27.4 | G/A | 0.77 (0.69, 0.86) | **1.9 x $10^{-6}$** | 0.50 (0.40, 0.62) | **2.3 x $10^{-10}$** | 0.86 (0.75, 0.98) | 0.025 | 0.81 (0.67, 0.97) | 0.025 |

OR: Odds Ratio. EA: Effect Allele for which OR is shown. OA: Other Allele. EAF: effect allele frequency (%) in the Icelandic sample. N: Number of TA cases. **Bold** indicates significant association after Bonferroni corrected $P = 0.05/68 = 0.00074$ (correcting for 17 markers and 4 phenotypes). *Strongest markers identified in the genome wide association studies of permanent tooth development[1], primary tooth development[16,17]. ‡When more than one variant has been associated with the trait at this locus in the previous studies and $r^2$ was below 0.2 between the markers, the variant with the lowest *P*-value in the previous studies were chosen.

**Appendix Table 3.** Genetic model analyses of the TA-associated variants.

| Phenotype/Marker | Locus/gene region (coding effect) | Position (hg38) | Alleles | MAF | $P_{mod}$* |
|---|---|---|---|---|---|
| **TOOTH AGENESIS** | | | | | *N* = 1,944 |
| rs4498834 | 1q32.1 *ASCL5*/*CACNA1S* | 201,111,170 | C/T | 43.0 | 0.29 |
| rs35822372 | 2p11.2 *FOXI3* | 88,438,931 | T/C | 21.0 | 0.61 |
| - | 2q13 *EDAR* (p.Arg420Trp) | 108,896,996 | A/G | 0.02 | NA |
| rs2034604 | 2q22.2 *ARHGAP15* | 143,201,176 | C/T | 47.3 | 0.93 |
| rs121908120 | 2q35 *WNT10A* (p.Phe228Ile) | 218,890,289 | A/T | 2.60 | **$5.2 \times 10^{-6}$** |
| rs121908119 | 2q35 *WNT10A* (p.Cys107Ter) | 218,882,368 | A/C | 0.14 | NA |
| rs371555610, rs529942527 | 8q21.13 *ZFHX4* | 76,604,644 | CTT/delCT T | 14.3 | 0.54 |
| **MANDIBULAR SECOND PREMOLARS** | | | | | *N* = 1,196 |
| rs917412 | 4q25 *LEF1* | 108,350,621 | T/C | 9.20 | 0.34 |
| **MAXILLARY SECOND PREMOLARS** | | | | | *N* = 600 |
| rs758468472 | 17q24.2 *NOL11* | 67,718,094 | G/T | 0.04 | NA |
| **MAXILLARY LATERAL INCISORS** | | | | | *N* = 600 |
| rs35956082 | 3p13 *FOXP1* | 71,414,748 | G/A | 25.2 | 0.70 |
| rs55846652 | Xq13.1 *EDA* | 69,564,858 | C/T | 32.8 | 0.24 |
| *Known OFC variant* | | | | | |
| rs5829552† | 2p24.2 *FAM49A* | 16,545,695 | TA/T | 22.5 | 0.34 |

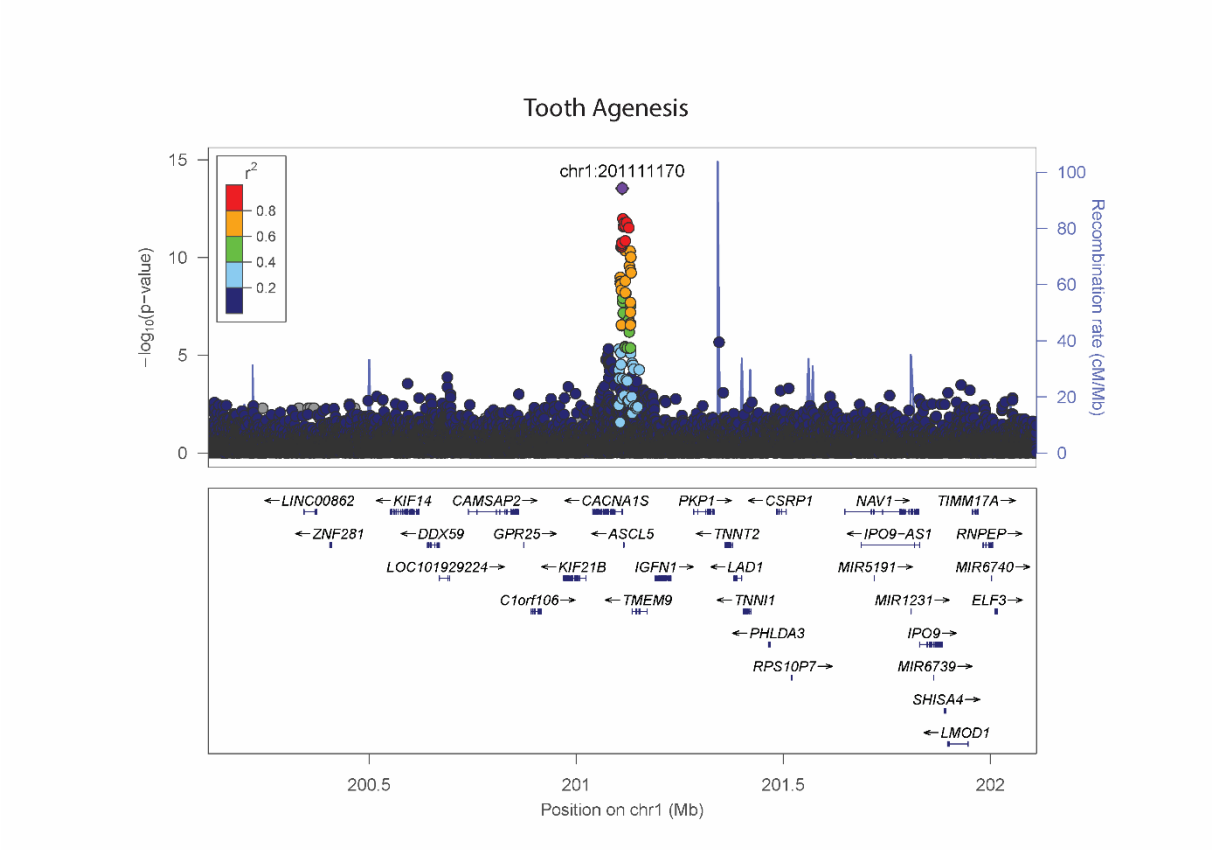
Alleles: minor/major allele. MAF: minor allele frequency (%). N: Number of TA cases.

*The genetic model analysis tested the multiplicative model versus the full model. Small *P* values indicate the multiplicative model should be rejected in favor of the full model. In the case of rs121908120, the ORs estimated for heterozygotes (3.0) and AA homozygotes (51.3) point to a recessive component. †The following rs-names are reported for this variant: rs5829552, rs397784935 and rs869032736.

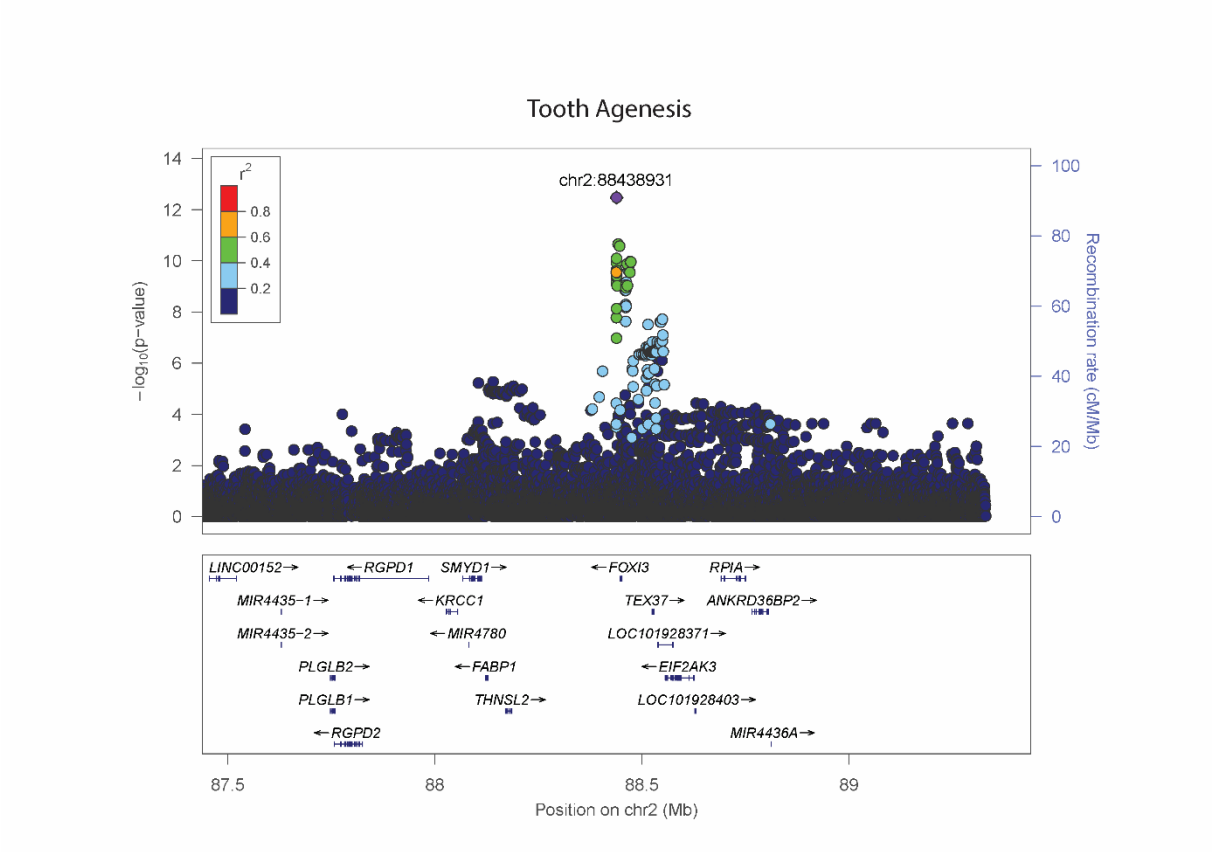
**Appendix Table 4.** INRICH gene-set enrichment analysis of the nine novel TA loci and *WNT10A*.

| Gene Ontology | #genes in set | #genes in test | *P* | *Pcorr* | Genes captured |
|---|---|---|---|---|---|
| GO:0042475\|\|odontogenesis_of_dentine-containing_tooth | 56 | 3 | 1.0 x 10-5 | **0.023** | *EDA,EDAR,LEF1* |
| GO:0061153\|\|trachea gland development | 2 | 2 | 1.0 x 10-5 | **0.023** | *EDA,LEF1* |
| GO:0010628\|\|positive_regulation_of_gene_expression | 248 | 4 | 2.0 x 10-5 | **0.031** | *EDA,EDAR,LEF1, WNT10A* |
| GO:0060662\|\|salivary_gland_cavitation | 5 | 2 | 2.0 x 10-5 | **0.031** | *EDA,EDAR* |
| GO:0042476\|\|odontogenesis | 28 | 2 | 3.0 x 10-5 | **0.042** | *LEF1,WNT10A* |
| GO:0042346\|\|positive_regulation_of_NF-kappaB_import_into_nucleus | 20 | 2 | 4.0 x 10-5 | **0.046** | *EDA,EDAR* |

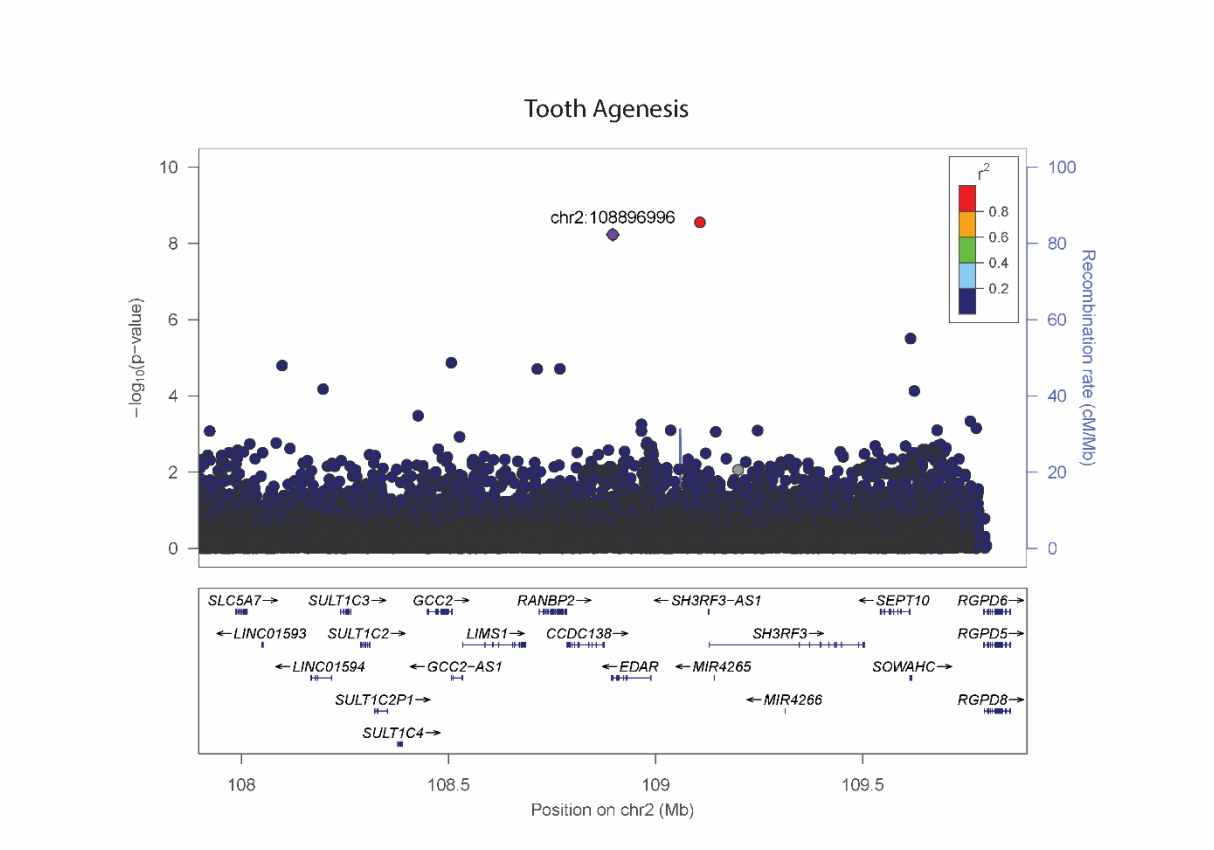
# Appendix Figures



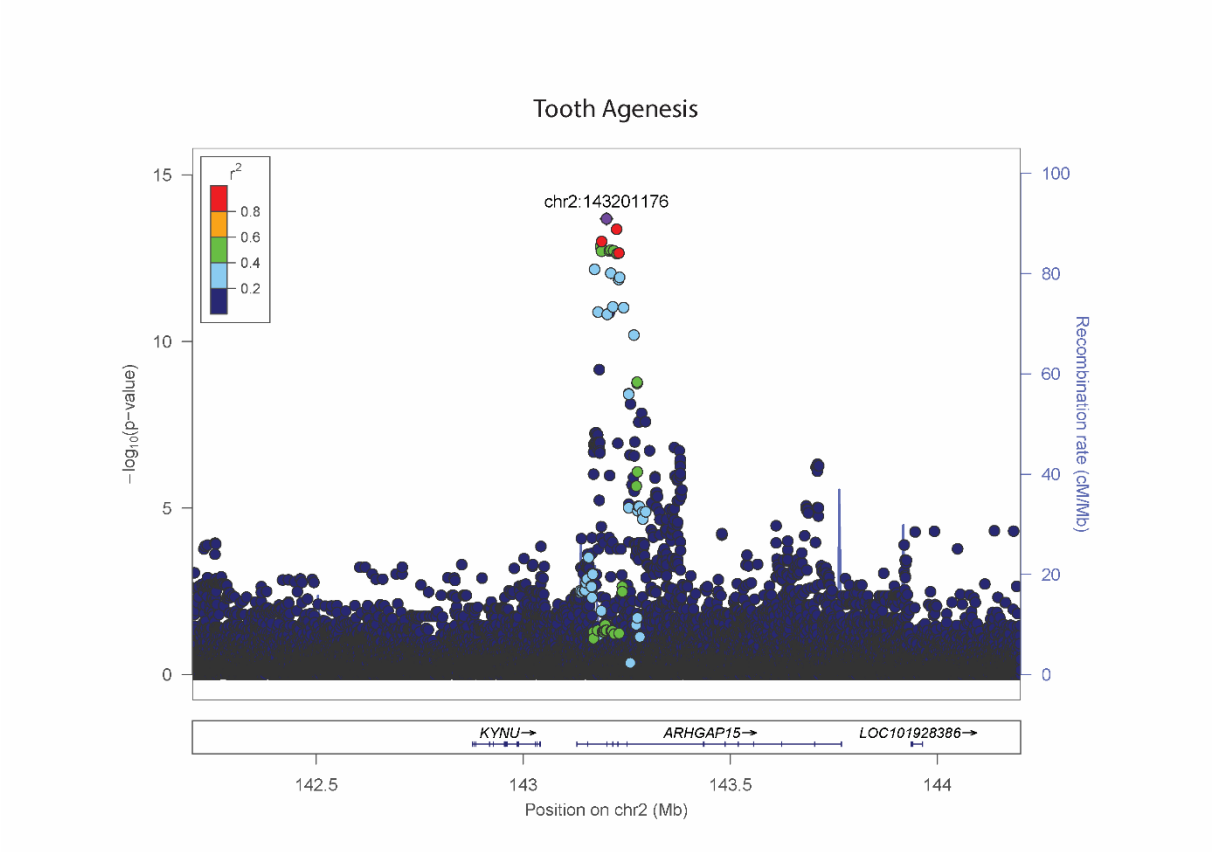

**Appendix Figure 1.** Regional association plots for rs4498834 (chr1:201111170) associated with Tooth Agenesis.



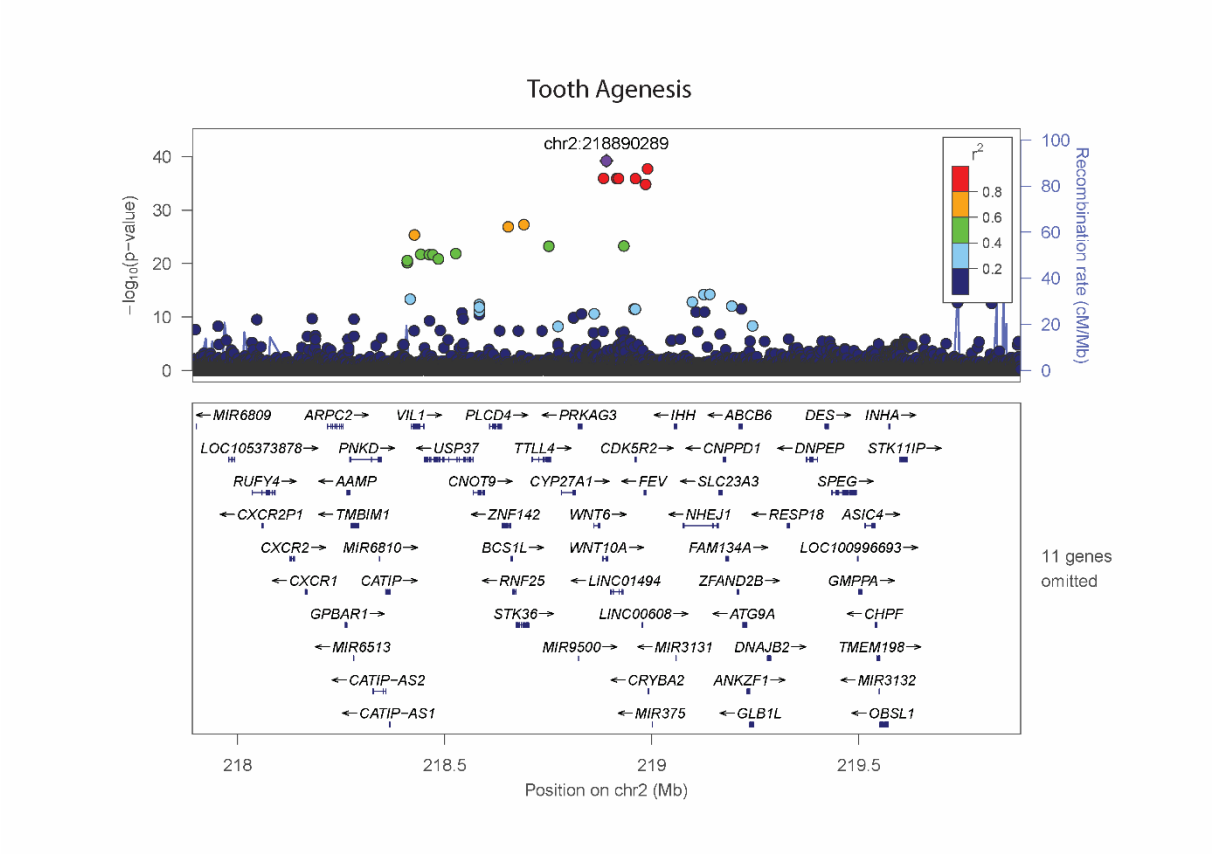

**Appendix Figure 2.** Regional association plots for rs35822372 (chr2:88438931) associated with Tooth Agenesis.



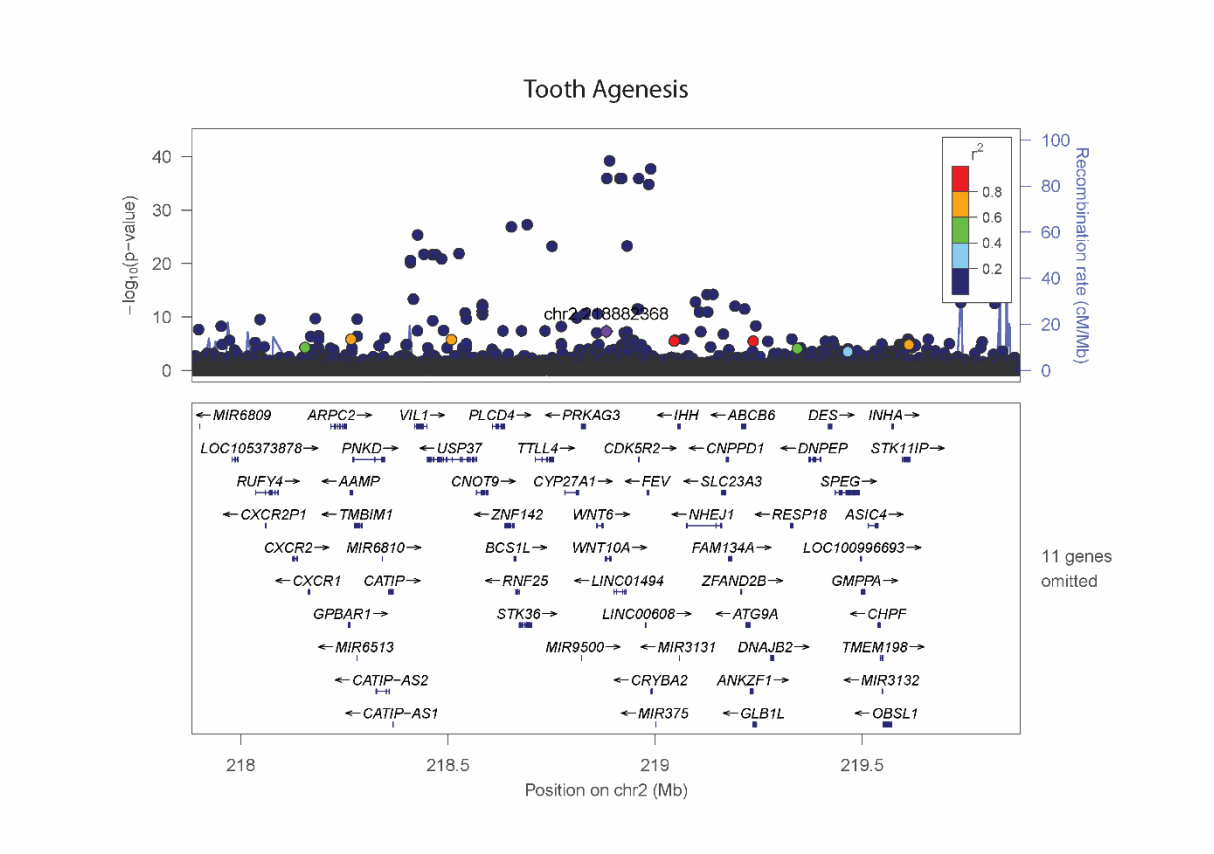

**Appendix Figure 3.** Regional association plots for *EDAR* (p.Arg420Trp, chr2:108896996) associated with Tooth Agenesis.



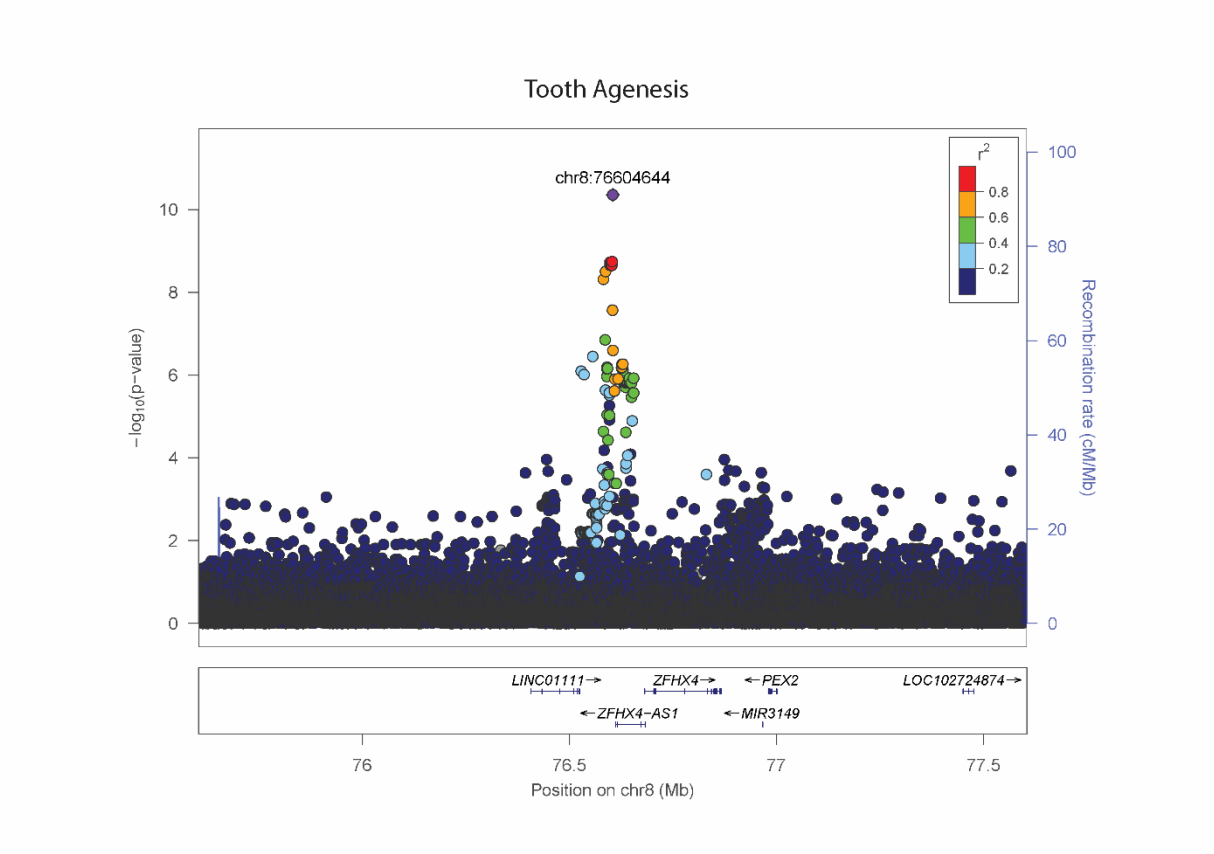

**Appendix Figure 4.** Regional association plots for rs2034604 (chr2:143201176) associated with Tooth Agenesis.



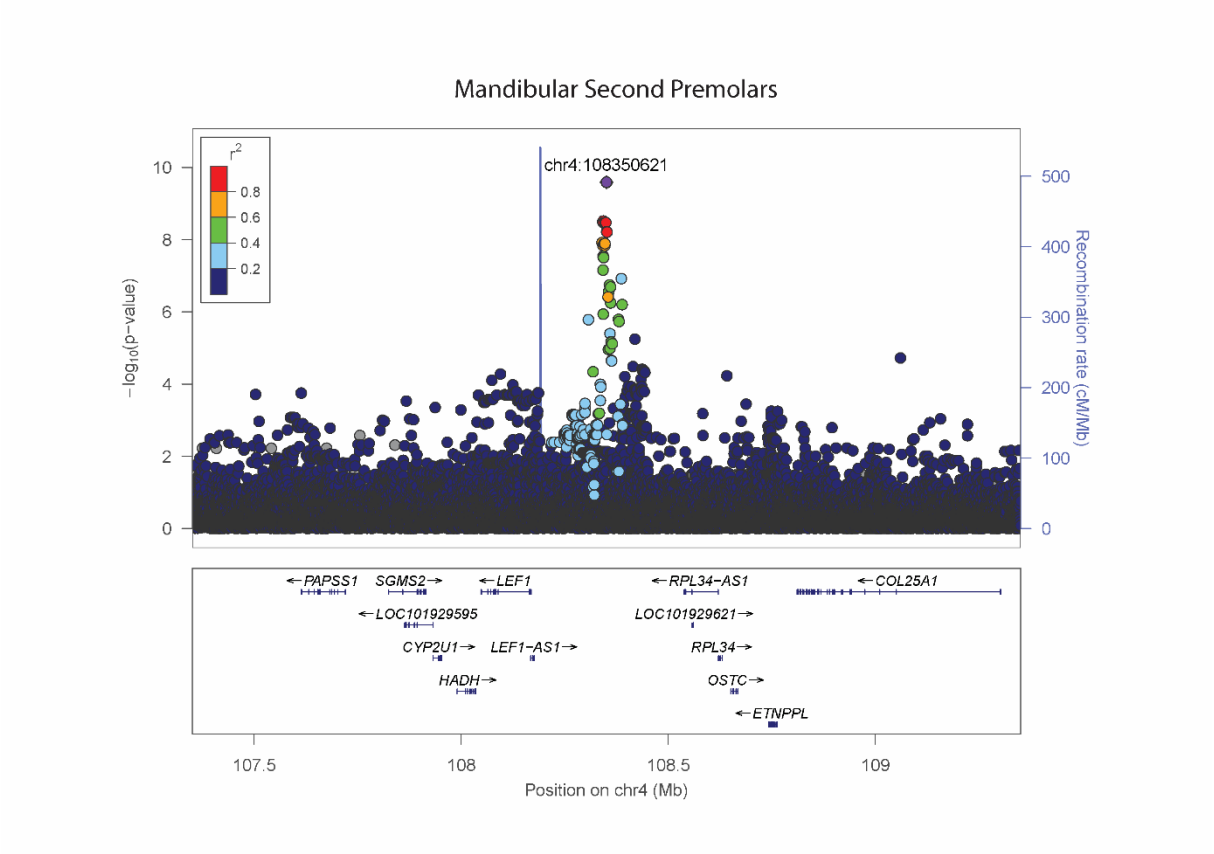

**Appendix Figure 5.** Regional association plots for rs121908120 (chr2:218890289) associated with Tooth Agenesis.



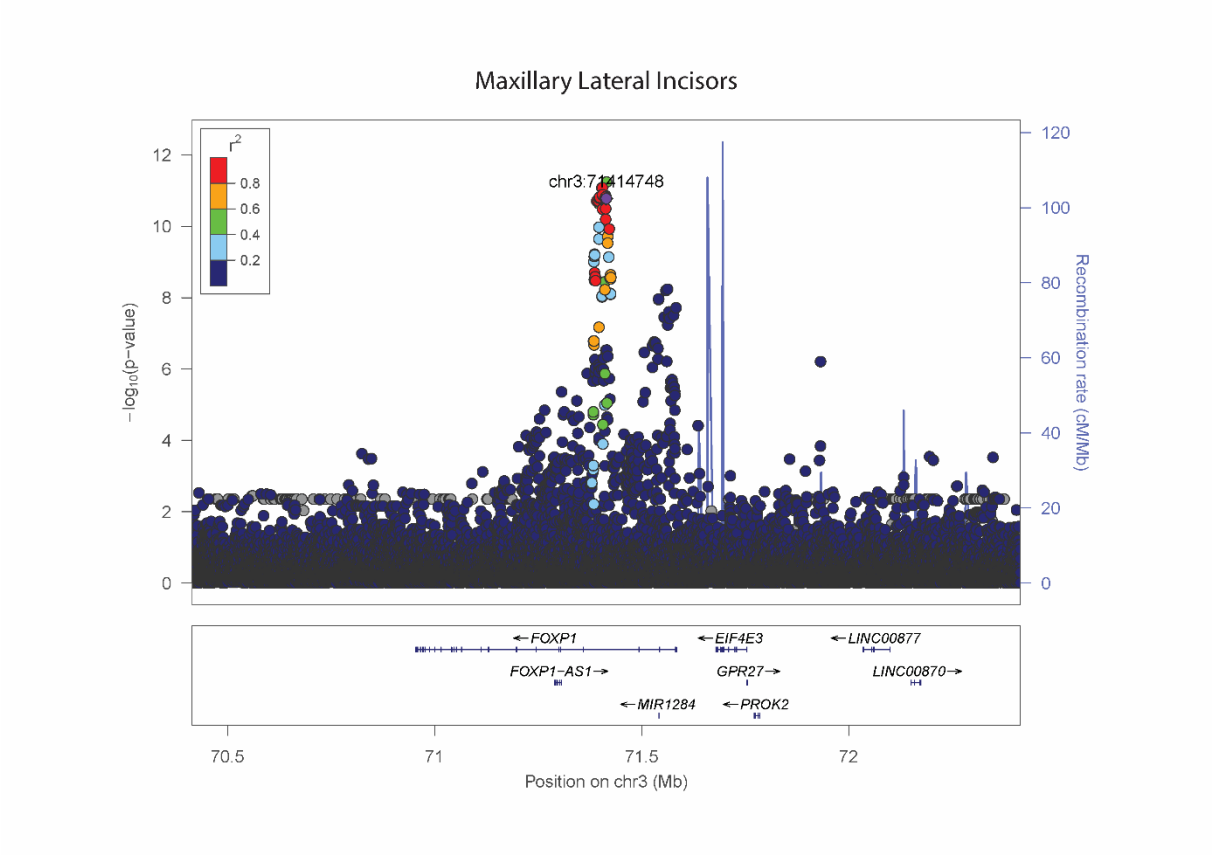

**Appendix Figure 6.** Regional association plots for rs121908119 (chr2:218882368) associated with Tooth Agenesis.



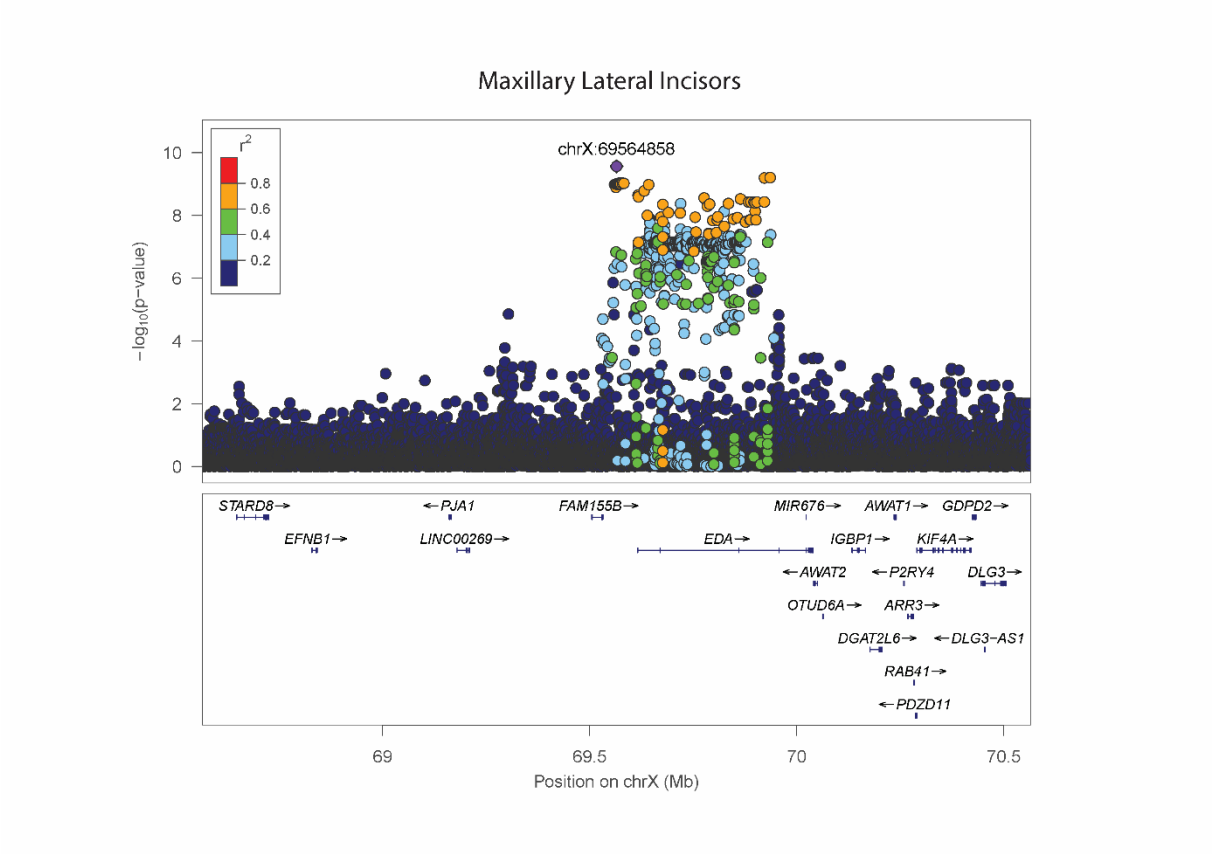

**Appendix Figure 7.** Regional association plots for rs371555610,rs529942527 (chr8:76604644) associated with Tooth Agenesis.



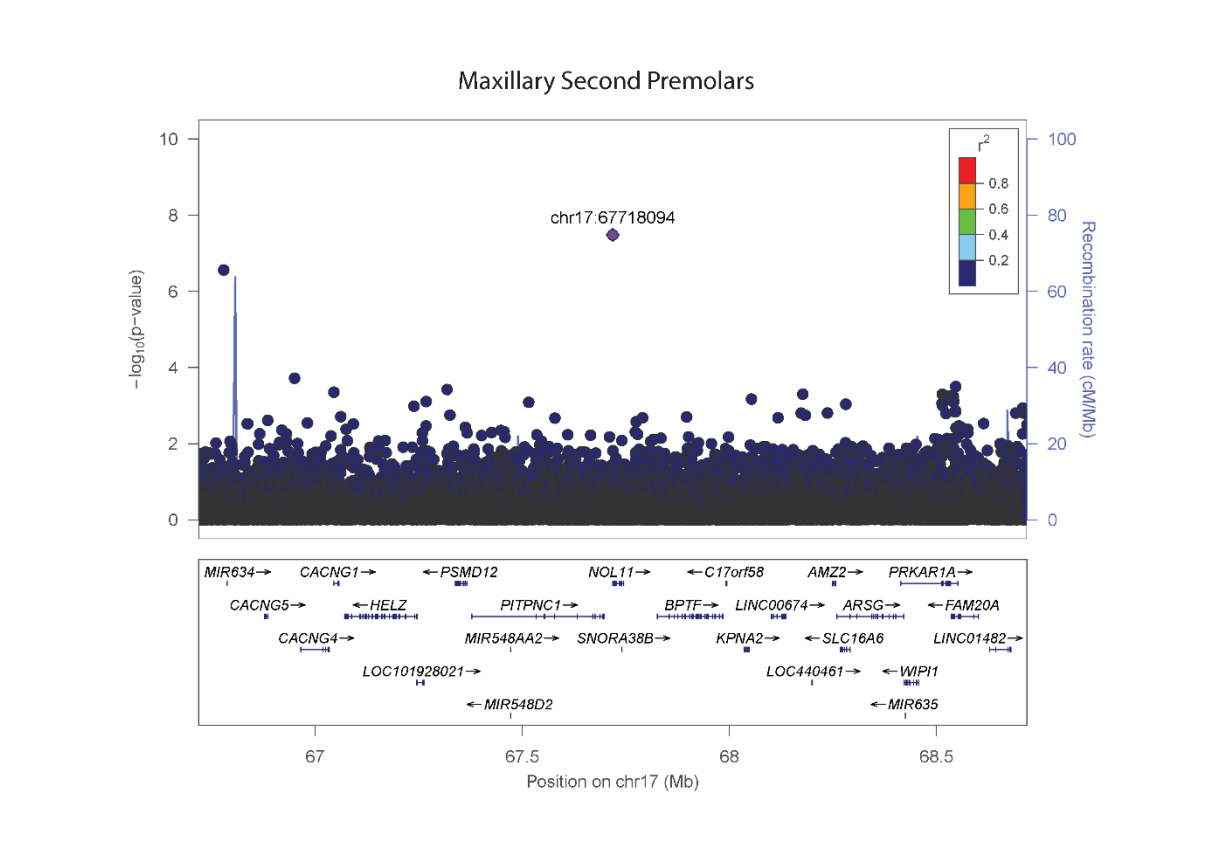

**Appendix Figure 8.** Regional association plots for rs917412 (chr4:108350621) associated with agenesis of Mandibular Second Premolar.




**Appendix Figure 9.** Regional association plots for rs35956082 (chr3:71414748) associated with Maxillary Lateral Incisors.




**Appendix Figure 10.** Regional association plots for rs55846652 (chrX:69564858) associated with Maxillary Lateral Incisors.




**Appendix Figure 11.** Regional association plots for rs758468472 (chr17:67718094) associated with Maxillary Second Premolars.

## References


1. Geller, F., Feenstra, B., Zhang, H. *et al.* Genome-wide association study identifies four loci associated with eruption of permanent teeth. *PLoS Genet* **7**, e1002275 (2011).
2. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
3. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
4. Helm, S. Recording system for the Danish Child Dental Health Services. *Community Dent Oral Epidemiol* **1**, 3-8 (1973).
5. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
6. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
7. Kruithof, C.J., Kooijman, M.N., van Duijn, C.M. *et al.* The Generation R Study: Biobank update 2015. *Eur J Epidemiol* **29**, 911-27 (2014).
8. Medina-Gomez, C., Felix, J.F., Estrada, K. *et al.* Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur J Epidemiol* **30**, 317-30 (2015).
9. Estrada, K., Abuseiris, A., Grosveld, F.G. *et al.* GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* **25**, 2750-2 (2009).
10. Howe, B.J., Cooper, M.E., Vieira, A.R. *et al.* Spectrum of Dental Phenotypes in Nonsyndromic Orofacial Clefting. *J Dent Res* **94**, 905-12 (2015).
11. Alves-Ferreira, M., Pinho, T., Sousa, A. *et al.* Identification of genetic risk factors for maxillary lateral incisor agenesis. *J Dent Res* **93**, 452-8 (2014).
12. von Elm, E., Altman, D.G., Egger, M. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology* **61**, 344-349 (2008).
13. Ludwig, K.U., Bohmer, A.C., Bowes, J. *et al.* Imputation of Orofacial Clefting Data Identifies Novel Risk Loci and Sheds Light on the Genetic Background of Cleft Lip +/- Cleft Palate and Cleft Palate Only. *Hum Mol Genet* (2017).
14. Yu, Y., Zuo, X., He, M. *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun* **8**, 14364 (2017).
15. Leslie, E.J., Carlson, J.C., Shaffer, J.R. *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet* **136**, 275-286 (2017).
16. Fatemifar, G., Hoggart, C.J., Paternoster, L. *et al.* Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. *Hum Mol Genet* **22**, 3807-17 (2013).
17. Pillas, D., Hoggart, C.J., Evans, D.M. *et al.* Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. *PLoS Genet* **6**, e1000856 (2010).
18. Lee, P.H., O'Dushlaine, C., Thomas, B. *et al.* INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797-9 (2012).